Depression rating scales for detection of major depression in people with dementia

Kirkham, Julia G.; Takwoingi, Yemisi; Quinn, Terry J.; Rapoport, Mark; Lanctôt, Krista L.; Maxwell, Colleen J.; Herrmann, Nathan; Gill, Sudeep S.; Rochon, Paula A.; Seitz, Dallas P.

DOI:
10.1002/14651858.CD012312

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Kirkham, JG, Takwoingi, Y, Quinn, TJ, Rapoport, M, Lanctôt, KL, Maxwell, CJ, Herrmann, N, Gill, SS, Rochon, PA & Seitz, DP 2016, 'Depression rating scales for detection of major depression in people with dementia', *Cochrane Database of Systematic Reviews*, vol. 2016, no. 8, CD012312.
https://doi.org/10.1002/14651858.CD012312

Cochrane Library

Cochrane Database of Systematic Reviews

Depression rating scales for detection of major depression in people with dementia (Protocol)

Kirkham JG, Takwoingi Y, Quinn TJ, Rapoport M, Lanctôt KL, Maxwell CJ, Herrmann N, Gill SS, Rochon PA, Seitz DP

Kirkham JG, Takwoingi Y, Quinn TJ, Rapoport M, Lanctôt KL, Maxwell CJ, Herrmann N, Gill SS, Rochon PA, Seitz DP.
Depression rating scales for detection of major depression in people with dementia.
Cochrane Database of Systematic Reviews 2016, Issue 8. Art. No.: CD012312.
DOI: 10.1002/14651858.CD012312.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Protocol]

Depression rating scales for detection of major depression in people with dementia

Julia G Kirkham[1], Yemisi Takwoingi[2], Terry J Quinn[3], Mark Rapoport[4], Krista L Lanctôt[5], Colleen J Maxwell[6], Nathan Herrmann[5], Sudeep S Gill[7], Paula A Rochon[8], Dallas P Seitz[1]

[1]Department of Psychiatry, Queen's University, Kingston, Canada. [2]Institute of Applied Health Research, University of Birmingham, Birmingham, UK. [3]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [4]Psychiatry, University of Toronto, Toronto, Canada. [5]Hurvitz Brain Sciences Research Program, Sunnybrook Research Institute, Toronto, Canada. [6]Schools of Pharmacy and Public Health and Health Systems, University of Waterloo, Waterloo, Canada. [7]Department of Medicine, Queen's University, Kingston, Canada. [8]Women's College Research Institute, Women's College Hospital, Toronto, Canada

Contact address: Dallas P Seitz, Department of Psychiatry, Queen's University, 752 King Street West, Kingston, ON, K7L 4X3, Canada. seitzd@providencecare.ca.

Editorial group: Cochrane Dementia and Cognitive Improvement Group.
Publication status and date: New, published in Issue 8, 2016.

Citation: Kirkham JG, Takwoingi Y, Quinn TJ, Rapoport M, Lanctôt KL, Maxwell CJ, Herrmann N, Gill SS, Rochon PA, Seitz DP. Depression rating scales for detection of major depression in people with dementia. *Cochrane Database of Systematic Reviews* 2016, Issue 8. Art. No.: CD012312. DOI: 10.1002/14651858.CD012312.

ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

To identify the accuracy of depression rating scales as screening tools for detecting DpD and compare the diagnostic accuracy of different depression rating scales for detecting MDD among adults with Alzheimer's disease and related forms of dementia.

To examine factors that may impact on the accuracy of depression rating scales that are used to diagnose depression. We will examine the reference standard used for verification of DpD, baseline prevalence of DpD in the study population, age of the underlying study population, gender of participants, type of dementia (any-cause dementia versus Alzheimer's disease), study setting (community or primary care setting, long-term care, tertiary care setting), and study country as potential sources of heterogeneity. We will also evaluate the effects of using different cut-points of individual depression rating scales on the diagnostic accuracy of the scales.

BACKGROUND

Target condition being diagnosed

Dementia currently affects approximately 35.6 million individuals worldwide, and the prevalence of dementia is anticipated to increase to 115 million by the year 2050 (Prince 2013). Depression has a bidirectional relationship with dementia: individuals with depression earlier in life have a two-fold increased risk of later developing dementia (Byers 2011; Byers 2012), and dementia is associated with six and a half times increased risk of developing significant symptoms of depression (Chen 1999; Djernes 2006; Starkstein 2005). As a result of this increased risk, an estimated 20% to 30% of people with Alzheimer's disease will develop major depressive disorder (MDD) (Castilla-Puentes 2010; Enache 2011; Lyketsos 1997; Weiner 2001). MDD in dementia (DpD) tends

to be chronic and persistent once it develops during the course of Alzheimer's disease (Aalten 2005; Steinberg 2004).

MDD is diagnosed clinically based on assessment by healthcare providers familiar with diagnostic criteria for MDD. The criteria for MDD in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) require the presence of a single or recurrent major depressive episode (American Psychiatric Association 2013). A major depressive episode consists of five or more symptoms of depression over a two-week period that represents a change from a previous level of functioning. One of the symptoms for a major depressive episode must be either depressed mood (either reported by the patient or observed by others) or loss of interest or enjoyment in most activities (either reported by the patient or others). Along with at least one of these two symptoms, a total of four other symptoms must be present including a significant change in weight or appetite, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue, feelings or worthlessness or guilt, difficulties with concentration, and recurrent thoughts of death or suicide. The symptoms of MDD must result in significant distress or difficulties in interpersonal or role functioning. These symptoms cannot be directly caused by a general medical condition or be caused by substances. The major depressive episode in MDD cannot occur as a part of another condition such as bipolar disorder or schizoaffective disorder, which can also have major depressive episodes. MDD can be further described in terms of its severity (mild, moderate, or severe) based on the number and severity of symptoms and resulting distress or dysfunction. MDD can also be specified as having the presence or absence of psychotic features, along with other specifiers for additional symptoms which may be present in the course of MDD such as anxiety or catatonia. The diagnosis of MDD in earlier versions of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (Text Revision) (DSM-IV-TR) are similar to those provided in the DSM-5 (American Psychiatric Association 2000), as are other criteria for MDD such as the International Classification of Diseases (ICD) (World Health Organization 1992). Specific criteria for the diagnosis of DpD have also been proposed (Olin 2002). These are similar to the criteria for MDD in populations without dementia except that only three symptoms of a major depressive episode are required (instead of five symptoms), with one of the symptoms being either depressed mood or loss of enjoyment in activities. Depressed mood or loss of enjoyment in activities can be identified either through patient report or caregiver or clinician observation of the patient. DpD can be diagnosed using either generic MDD criteria or criteria specific for DpD.

Clinical practice guidelines, in Canadian Coalition for Seniors' Mental Health 2006 and Herrmann 2008, and reviews, in Kales 2014, Pieper 2011, Seitz 2012, Seitz 2013, and Sink 2005, recommend that individuals with dementia be assessed for MDD, and the identification and treatment of DpD is an important part of providing care for individuals with dementia for several reasons. Most individuals with DpD are not treated, often due to underdiagnosis of DpD (Payne 2002; van Asch 2013), and untreated DpD is associated with increased mortality (Geerlings 1999), accelerated cognitive decline (Hargrave 2000), earlier nursing home placement (Dorenlot 2005), and decreased quality of life (Winter 2011). Underdiagnosis of DpD can be associated with increased use of medications such as benzodiazepines and antipsychotics (Evers 2002; Volicer 2011), while overdiagnosis can lead to the inappropriate use of antidepressants (Berman 2012). When DpD is accurately identified, effective nonpharmacological interventions for depression can be implemented (McSweeney 2012; Orgeta 2014; Teri 1997; Teri 2003; Teri 2005). The efficacy of pharmacological interventions for DpD is unclear at this time (Banerjee 2011; Nelson 2011); further studies are underway to evaluate the use of medications for treating DpD.

Index test(s)

A number of depression rating scales have been investigated as possible tools to diagnose DpD (Alexopoulos 1988; Hancock 2009; Knapskog 2012; Muller-Thomsen 2005; Schreiner 2003; Sunderland 1988). However, these scales for DpD differ in content and mode of administration, and their comparative diagnostic accuracy has not been synthesized to date. We have provided a table summarizing key characteristics of these rating scales in Appendix 1.

Clinical pathway

Depression can be difficult to identify in individuals with dementia, as they may not be able to communicate symptoms and because symptoms of dementia, such as apathy, can mimic depression. As a result, DpD can be both underdiagnosed (that is false negatives), Volicer 2011, and overdiagnosed (that is false positives), Starkstein 2005, in clinical settings. The majority of individuals with dementia will first be assessed by primary care providers such as nurses or family physicians. People with dementia may present for evaluation of potential depression with symptoms such as a reported decrease in mood or loss of enjoyment in activities, or with non-specific changes in sleep or appetite. Individuals with milder severity of dementia may be able to self report symptoms of depression similar to individuals without dementia. However, individuals with more advanced dementia may not be able to report these symptoms, and assessments incorporating clinician and caregiver observations may be required to identify depression in this population. In both community and long-term care settings, depression rating scales may be used by primary care providers or non-mental health clinicians to screen for DpD. Those individuals who screen positive for possible DpD on these rating scales may then go on to have further evaluation either by primary care providers or be referred to specialists such as psychologists or psychiatrists for further evaluation and confirmation of the diagnosis.

The reference standard for the diagnosis of DpD is clinical examination and diagnosis by geriatric psychiatrists or psychologists using standardized diagnostic clinical criteria (American Psychiatric Association 2000; Olin 2002). In some settings, screening positive on a depression rating scale may be used to initiate a referral to mental health specialists, or primary care providers may decide to initiate non-pharmacological or pharmacological treatment for DpD based on the results of the depression rating scale and clinical evaluation without referral to other services.

Alternative test(s)

Diagnosis of depression in dementia and other contexts relies on clinical evaluation, and to date no validated alternative tests exist.

Rationale

There are gaps in the current knowledge base on practical strategies for the diagnosis of DpD in routine clinical settings. The optimal cut-points for diagnosing DpD on various depression rating scales have not been consistently defined, and the comparative accuracy of these different rating scales has not been described. In addition, the impact of factors such as patient characteristics, reference standard used, setting (community/primary care, long-term care, tertiary care), type and severity of dementia, and methods of scale administration on the accuracy of depression scales remains unclear. To date, there are no published reviews on the accuracy of depression rating scales for the diagnosis of DpD. Our search of the PROSPERO registry and correspondence with the Cochrane Dementia and Cognitive Improvement Group have not identified any registered titles or protocols on this topic. To our knowledge, our review will be the first to answer these important questions regarding the accuracy of depression rating scales for diagnosing DpD.

OBJECTIVES

To identify the accuracy of depression rating scales as screening tools for detecting DpD and compare the diagnostic accuracy of different depression rating scales for detecting MDD among adults with Alzheimer's disease and related forms of dementia.

Secondary objectives

To examine factors that may impact on the accuracy of depression rating scales that are used to diagnose depression. We will examine the reference standard used for verification of DpD, baseline prevalence of DpD in the study population, age of the underlying study population, gender of participants, type of dementia (any-cause dementia versus Alzheimer's disease), study setting (community or primary care setting, long-term care, tertiary care setting), and study country as potential sources of heterogeneity. We will also evaluate the effects of using different cut-points of individual depression rating scales on the diagnostic accuracy of the scales.

METHODS

Criteria for considering studies for this review

Types of studies

We will follow general recommendations for the completion of diagnostic test accuracy reviews (Cochrane 2013; Davis 2013). We will focus our review on studies that assess all participants in the study sample with both the index tests and reference standards for DpD. We will include cross-sectional studies using any depression rating scale in undifferentiated study populations diagnosed with dementia.

Participants

Participants are individuals with a diagnosis of dementia using standardized diagnostic criteria (for example DSM-IV-TR (American Psychiatric Association 2000), DSM-5 (American Psychiatric Association 2013), ICD 9 or 10 (World Health Organization 1992), or National Institute of Neurological Disorders and Stroke - Alzheimer's Disease and Related Disorders Association for Alzheimer's disease (McKhann 1984)). We will include studies of participants with dementia due to Alzheimer's disease, vascular dementia (Roman 1993), mixed vascular and Alzheimer's disease dementia, and dementia with Lewy bodies (McKeith 2005), as these are the most common forms of dementia and there is significant overlap in pathology between these types of dementia (Jellinger 2006). We will exclude less common forms of dementias, such as Parkinson's disease dementia, frontotemporal dementia, progressive supranuclear palsy, multiple system atrophy, or dementia due to general medical conditions such as HIV or Huntington's disease. In clinical practice, depression rating scales may be used to diagnose DpD or as a screening tool to identify individuals who if screened positive on the index test would go on for further evaluation and confirmation of DpD using the reference standard. We will exclude studies that only conducted the reference standard on the subgroup of the study population that screened positive for DpD on the index tests.

Index tests

The index tests are depression rating scales, including self report or interviewer-based scales, which have been used to diagnose DpD. We will not a priori restrict our review to any specific scales. Preliminary literature searches have identified several scales which have been used to diagnose DpD, including several generic depression rating scales (for example Hamilton Depression Rating Scale (Lichtenberg 1992; Naarding 2002), Montgomery-Åsberg Depression Rating Scale (da Gloria 2012; Knapskog 2012; Leontjevas 2012), Geriatric Depression Scale (Bedard 2003; Debruyne 2009; Feher 1992; Korner 2006; Lichtenberg 1992; Muller-Thomsen 2005; Schreiner 2003; Snow 2005), Nursing Observations Scale for Geriatric Patients (Muller-Thomsen 2005), Patient Health Questionnaire-9 (Hancock 2009), Beck Depression Inventory (Beck 1961)) as well as rating scales specific to DpD (for example Cornell Depression in Dementia Scale (Alexopoulos 1988; Barca 2010; da Gloria 2012; Knapskog 2012; Korner 2006; Leontjevas 2012; Muller-Thomsen 2005; Schreiner 2003) and Dementia Mood Assessment Scale (Sunderland 1988)). These scales are continuous scales that rate the severity or frequency of symptoms of depression. Various cut-points on these scales have been used as a threshold to categorize individuals as being positive or negative for DpD (da Gloria 2012; Debruyne 2009; Knapskog 2012; Leontjevas 2012). We have provided a summary of the characteristics of these scales in Appendix 1.

Target conditions

Diagnostic criteria for major depressive disorder require the presence of either a depressed mood or a decreased interest in activities for the majority of a two-week time period. Along with at least one of these core symptoms, individuals must also experience at least four additional symptoms including: changes in sleep, psychomotor activities, appetite changes, feelings of guilt or worthlessness, difficulties with concentration, decreased energy, or thoughts of suicide or death. The individual must experience significant distress or dysfunction, and these symptoms must not be better accounted for by the effects of a substance or a medical condition.

Reference standards

Major depressive disorder is a clinical diagnosis. The reference standard for the diagnosis of DpD includes several depression diagnostic criteria administered by a mental health professional. The most commonly used depression diagnostic criteria are those contained in the DSM. The criteria used in the various iterations of DSM (for example III, IV-TR, 5) are relatively similar. These diagnostic criteria for major depressive disorder can be used for individuals with and without underlying dementia, and their inter-rater reliability is moderate to high (Brown 2001; Keller 1995; Regier 1994; Spitzer 1979). Specific criteria for the diagnosis of DpD have also been developed (Olin 2002). These criteria are similar to the DSM-IV diagnostic criteria except that the core criteria of a sad or depressed mood or reported lack of enjoyment can be based on observations of the patient and fewer associated symptoms are required (three total symptoms instead of five) (Olin 2002). Other depression diagnostic criteria, including the ICD 9 or 10, are similar to those of the DSM.

Search methods for identification of studies

Electronic searches

We will search MEDLINE (Ovid SP), EMBASE (Ovid SP), CINAHL (EBSCOhost), LILACS (BIREME), BIOSIS Previews (Thomson Reuters Web of Science) and Web of Science Core Collection, including Conference Proceedings Citation Index (Thomson Reuters Web of Science). See Appendix 2 for a proposed MEDLINE search strategy plus an additional narrative on the search strategy and process. We will apply no restrictions based on language of publication, and potentially elgiible studies in languages other than English will be translated.

The information specialist for the Cochrane Dementia and Cognitive Improvement Group (Anna Noel-Storr) will perform the initial searches.

Searching other resources

We will search the reference lists of relevant studies and reviews to identify potentially relevant studies that may have been missed by the electronic searches (Greenhalgh 2005; Horsely 2011). We will use the related articles feature in electronic databases to identify relevant studies. We will contact authors of included studies for unpublished data that may be included in the analysis and to identify unpublished study reports.
We will not perform handsearching, as there is little published evidence of the benefits of handsearching for reports of diagnostic test accuracy studies (Beynon 2013; Whiting 2011b).

Data collection and analysis

Selection of studies

The inclusion criteria for studies in this review are the following.

1. Study population with a diagnosis of dementia using standardized criteria (e.g. DSM-IV-TR (American Psychiatric Association 2000), DSM-5 (American Psychiatric Association 2013), ICD 9 or 10 (World Health Organization 1992), National Institute of Neurological Disorders and Stroke - Alzheimer's Disease criteria (McKhann 1984)), vascular

dementia (Roman 1993), mixed Alzheimer's disease and vascular dementia or dementia with Lewy bodies (McKeith 2005).

2. Reference standard diagnosis of major depressive disorder of any severity or subtype diagnosed by a psychiatrist or other qualified mental health practitioner according to either generic diagnostic criteria (e.g. DSM-IV-TR, American Psychiatric Association 2000) or specific diagnostic criteria for DpD (Olin 2002).

3. Index test including any depression rating scale used to diagnose DpD; these scales may be generic depression rating scales, in Knapskog 2012, Lichtenberg 1992, Muller-Thomsen 2005, and Schreiner 2003, or depression rating scales specific to DpD (Alexopoulos 1988; Sunderland 1988); index and reference test administered within two weeks.

We will exclude the following studies.

1. Case-control studies.

2. Studies where there is insufficient information to recreate the 2x2 table of the number of true positives, false positives, false negatives, and true negatives.

3. Studies that report depressive disorders other than major depression.

4. Studies where participants have cognitive impairment that does not meet the criteria for dementia.

Data extraction and management

Two review authors will independently extract information from studies meeting the inclusion criteria. We will pilot, refine, and then use a data extraction form (Appendix 3) to record information from each included study. We will extract the following data.

1. Study design and setting.

2. Characteristics of the study population such as median/mean age and range, sex, place of residence of participants, educational status, cognitive testing (e.g. Mini Mental State Exam score (Folstein 1975)), criteria used for diagnosis of dementia, type of dementia (e.g. Alzheimer's disease or all-cause dementia), severity of underlying dementia (e.g. Clinical Dementia Rating Scale, Hughes 1982, or Global Deterioration Scale, Reisberg 1982, scores), prior history of depression and other mental disorders.

3. Prevalence of DpD.

4. Depression rating scales along with the characteristics of individuals administering each scale. We will also record the cut-points used to identify participants as having a positive result for DpD.

5. Reference standard criteria for the diagnosis of DpD along with characteristics of the individuals administering the reference standard.

6. Number of true positives, true negatives, false positives, and false negatives or summary statistics that will enable their derivation. We will use these values to create 2x2 tables for each rating scale and enter the data into Review Manager version 5.3 (Review Manager 2014).

Assessment of methodological quality

We will assess the quality of studies using the QUADAS-2 tool, which is recommended by The Cochrane Collaboration for appraising the methodological quality of diagnostic accuracy studies (Whiting 2011a). The QUADAS-2 tool assesses risk of bias in four domains: patient selection, index test, reference standard, and participant flow and timing. Each domain is rated as being at high, low, or unclear risk of bias based on responses to a series of signalling questions. In addition to risk of bias, the tool also considers issues of applicability related to patient selection, index test, and reference standard. We have tailored the tool to our review question and developed guidance on how to assess each signalling question as shown in Appendix 4. We will summarize results of the quality assessment in tables or figures or both. We will consider the quality of the evidence when interpreting the findings of the review and drawing conclusions.

Statistical analysis and data synthesis

We will analyse studies separately based on study setting: community or primary care; long-term care or nursing homes; or tertiary care. We will also separately analyze studies that used generic diagnostic criteria for MDD and those that used reference standard criteria that are specific to DpD (Olin 2002).

We will create coupled forest plots of sensitivity and specificity for each rating scale at one or more cut-points. Study-specific estimates of sensitivity and specificity will also be plotted in receiver operating characteristic (ROC) space. We will use these plots for preliminary investigations of the data and to visually explore heterogeneity.

For each scale, if most studies report common cut-points, we will perform meta-analysis by using the bivariate model to estimate summary points (Chu 2006; Reitsma 2005). Alternatively, if studies report different cut-points, we will use the hierarchical summary ROC (HSROC) model to estimate SROC curves (Rutter 2001). Since sensitivity and specificty are useful summary measures for interpreting the consequences of test errors (false negatives and false positives), we will quantify test performance from a SROC curve by estimating sensitivity at points on the curve that correspond to the lower quartile, median and upper quartile of the specificities observed in the studies included in the meta-analysis. If a study provides 2x2 tables for more than one cut-point, one table will be selected at random for the HSROC analysis. However, if several studies report data at several cut-points, we will consider using methods that allow for such data (Dukic 2003; Hamza 2009; Riley 2015). Given the complexity of hierarchical models, where few studies are available we will simplify the models by removing model parameters as has been recommended (Takwoingi 2015).

To formally compare the accuracy of the scales, we will add a covariate indicating test type to a hierarchical model (bivariate

or HSROC model) to assess differences in the accuracy of the scales. We will assess the statistical significance of differences in test performance using likelihood ratio tests to compare models with and without the covariate terms. As comparative studies that directly compare test accuracy are not often available (Takwoingi 2013), we plan to perform both indirect (include all available studies) and direct test comparisons. Direct comparisons will be performed as pairwise comparisons of studies that have compared two scales head-to-head in the same study population.

We will use the NLMIXED procedure in the SAS software (version 9.4; SAS Institute, Cary, NC, USA) to fit HSROC models or the meqrlogit command in Stata 14 (StataCorp, College Station, TX, USA) to fit bivariate models.

Investigations of heterogeneity

We will perform investigations of heterogeneity using visual examination of forest plots and summary ROC plots in the first instance. Where there is sufficient data, we will perform meta-regression by including each factor specified in our secondary objectives as a covariate in a hierarchical model to determine its effect on test accuracy.

Sensitivity analyses

We will repeat the primary meta-analyses excluding studies judged as having a high risk of bias on the QUADAS-2 assessment for the index test and reference standard domains.

Assessment of reporting bias

Little is known about the determinants and extent of publication bias for test accuracy studies. Traditionally used methods for assessing publication bias are not recommended for test accuracy reviews. The most suitable approach for test accuracy reviews has low power when there is heterogeneity (Deeks 2005). Since heterogeneity is expected in test accuracy reviews and is likely in this review, we will not assess publication bias using any of the tests of funnel plot asymmetry (Deeks 2005; van Enst 2014).

ACKNOWLEDGEMENTS

The review authors would like to thank Sue Marcus and Dr. Rupert McShane from the Cochrane Cognitive Improvement and Dementia Group for providing preliminary feedback on this proposal. The review authors would also like to thank Anna Noel-Storr for assisting with drafting the electronic search strategy.

REFERENCES

Additional references

Aalten 2005
Aalten P, de Vugt ME, Jaspers N, Jolles J, Verhey FR. The course of neuropsychiatric symptoms in dementia. Part I: findings from the two-year longitudinal Maasbed study. *International Journal of Geriatric Psychiatry* 2005;**20**(6): 523–30.

Alexopoulos 1988
Alexopoulos GS, Abrams, RC, Young RC, Shamoian CA. Cornell scale for depression in dementia. *Biological Psychiatry* 1988;**23**(3):271–84.

American Psychiatric Association 2000
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disoders, Text Revision*. 4th Edition. Washington, DC: American Psychiatric Association, 2000.

American Psychiatric Association 2013
American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th Edition. Washington, DC: American Psychiatric Association, 2013.

Banerjee 2011
Banerjee S, Hellier J, Dewey M, Romeo R, Ballard C, Baldwin R, et al. Sertraline or mirtazapine for depression in dementia (HTA-SADD): a randomized, multicentre, double-blind, placebo-controlled trial. *The Lancet* 2011; **378**(9789):403–11.

Barca 2010
Barca ML, Engedal K, Selbaek G. A reliability and validity study of the Cornell Scale among elderly inpatients, using various clinical criteria. *Dementia and Geriatric Cognitive Disorders* 2010;**29**(5):438–47.

Beck 1961
Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of General Psychiatry* 1961;**4**:561–71.

Bedard 2003
Bedard M, Squire L, Minthorn-Biggs MB, Molloy W, Dubois S, O'Donnell M, et al. Validity of self-reports in dementia research: the Geriatric Depression Scale. *Clinical Gerontologist* 2003;**26**(3-4):155–63.

Berman 2012
Berman K, Brodaty H, Withall A, Seeher K. Pharmacologic treatment of apathy in dementia. *American Journal of Geriatric Psychiatry* 2012;**20**(2):104–22.

Beynon 2013
Beynon R, Leeflang MM, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. *Cochrane*

Database of Systematic Reviews 2013, Issue 9. [DOI: 10.1002/14651858.MR000022.pub3]

Brown 2001

Brown TA, Di Nardo Pa, Lehman CL, Campbell LA. Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. *Journal of Abnormal Psychology* 2001;**110**(1):49–58.

Byers 2011

Byers AL, Yaffe K. Depression and risk of developing dementia. *Nature Reviews* 2011;**7**(6):323–31.

Byers 2012

Byers A, Covinsky K, Barnes D, Yaffe K. Dysthymia and depression increase risk of dementia and mortality among older veterans. *American Journal of Geriatric Psychiatry* 2012;**20**(8):664–72.

Canadian Coalition for Seniors' Mental Health 2006

Canadian Coalition for Seniors' Mental Health. *The Assessment and Treatment of Mental Health Issues in Long Term Care Homes*. Toronto, ON: Canadian Coalition for Seniors' Mental Health, 2006.

Castilla-Puentes 2010

Castilla-Puentes RC, Habeych MR. Subtypes of depression among patients with Alzheimer's disease and other dementias. *Alzheimer's & Dementia* 2010;**6**(1):63–9.

Chen 1999

Chen P, Ganguli M, Mulsant B, DeKosky S. The temporal relationship between depressive symptoms and dementia: a community-based prospective study. *Archives of General Psychiatry* 1999;**56**(3):261–6.

Chu 2006

Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006;**59**: 1331–2.

Cochrane 2013

The Cochrane Collaboration Diagnostic Test Accuracy Working Group. *Handbook for Diagnostic Test Accuracy*. The Cochrane Collaboration, 2013.

da Gloria 2012

da Gloria PM, Coutinho ESF, Almeida C, Barca ML, Knapskog AB, Engedal K, et al. Validation of Montgomery-Asberg Rating Scale and Cornell Scale for Depression in dementia in Brazilian elderly patients. *International Psychgeriatrics* 2012;**24**(8):1291–8.

Davis 2013

Davis DH, Creavin ST, Noel-Storr A, Quinn TJ, Smailagic N, Hyde C, et al. Neuropsychological tests for the diagnosis of Alzheimer's disease dementia and other dementias: a generic protocol for cross-sectional and delayed-verification studies. *Cochrane Database Systematic Reviews* 2013, Issue 3. [DOI: 10.1002/14651858.CD010460]

Debruyne 2009

Debruyne H, Van Buggenhout M, Le Bastard N, Aries M, Audenaert K, De Deyn PP, et al. Is the Geriatric Depression Scale a reliable screening tool for depressive symptoms in elderly patients with cogntive impairment. *International Journal of Geriatric Psychiatry* 2009;**24**(6):556–62.

Deeks 2005

Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005;**58**(9):882–93.

Djernes 2006

Djernes J. Prevalence and predictors of depression in populations of elderly: a review. *Acta Psychiatrica Scandinavica* 2006;**113**(5):372–87.

Dorenlot 2005

Dorenlot P, Harboun M, Bige V, Henrard J, Ankri J. Major depression as a risk factor for early institutionalization of dementia patients living in the community. *International Journal of Geriatric Psychiatry* 2005;**20**(5):471–8.

Dukic 2003

Dukic V, Gatsonis C. Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics* 2003;**59**(4):936–46.

Enache 2011

Enache D, Winblad B, Aarsland D. Depression in dementia: epidemiology, mechanisms, and treatment. *Current Opinion in Psychiatry* 2011;**24**(6):461–72.

Evers 2002

Evers M, Samuels S, Lantz M, Khan K, Brickman A, Marin D. The prevalence, diagnosis and treatment of depression in dementia patients in chronic care facilities in the last six months of life. *International Journal of Geriatric Psychiatry* 2002;**17**(5):464–72.

Feher 1992

Feher EP, Larrabee GJ, Crook TH. Factors attenuating the validity of the Geriatric Depression Scale in a dementia population. *Journal of the American Geriatrics Society* 1992; **40**(9):906–9.

Folstein 1975

Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;**12**(3):189–98.

Geerlings 1999

Geerlings M, Deeg D, Penninx B, Schmand B, Jonker C, Bouter LM, et al. Cognitive reserve and mortality in dementia: the role of cognition, functional ability and depression. *Psychological Medicine* 1999;**29**(5):1219–26.

Greenhalgh 2005

Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ* 2005;**331**(7524):1064–5.

Hamza 2009

Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Medical Research Methodology* 2009;**9**(1):73.

Hancock 2009
Hancock P, Larner A. Clinical utility of the Patient Health Questionnaire-9 (PHQ-9) in memory clinics. *International Journal of Psychiatry in Clinical Practice* 2009;**13**(3):188–91.

Hargrave 2000
Hargrave R, Reed B, Mungas D. Depressive syndromes and functional disability in dementia. *Journal of Geriatric Psychiatry and Neurology* 2000;**13**(2):72–7.

Herrmann 2008
Herrmann N, Gauthier S. Diagnosis and treatment of dementia: 6. Management of severe Alzheimer Disease. *Canadian Medical Association Journal* 2008;**179**(12): 1279–87.

Horsely 2011
Horsely T, Dingwall O, Sampson M. Checking reference lists to find additional studies for systematic reviews. *Cochrane Database of Systematic Reviews* 2011, Issue 8. [DOI: 10.1002/14651858.MR000026.pub2]

Hughes 1982
Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *The British Journal of Psychiatry* 1982;**140**(6):566–72.

Jellinger 2006
Jellinger KA. Clinicopathological analysis of dementia disorders in the elderly - an update. *Journal of Alzheimer's Disease* 2006;**9**(3 Suppl):61–70.

Kales 2014
Kales HC, Gitlin LN, Lyketsos CG. Management of neuropsychiatric symptoms of dementia in clinical settings: recommendations from a multidisciplinary expert panel. *Journal of the American Geriatrics Society* 2014;**62**(4):762–9.

Keller 1995
Keller MB, Kein DN, Hirschfeld RM, Kocsis JH, McCullough JP, Miller I. Results of the DSM-IV mood disorders field trial. *American Journal of Psychiatry* 1995;**1** (52):843–9.

Knapskog 2012
Knapskog AB, Barca ML, Engedal K. A comparison of the validity of the Cornell Scale and the MADRS in detecting depression among memory clinic patients. *Dementia and Geriatric Cognitive Disorders* 2012;**32**(4):287–94.

Korner 2006
Korner A, Lauritzen L, Abelskov K, Gulmann N, Brodersen A, Wedervang-Jensen T, et al. The Geriatric Depression Scale and the Cornell Scale for Depression in dementia: a validity study. *Nordic Journal of Psychiatry* 2006;**60**(5): 360–4.

Leontjevas 2012
Leontjevas R, Gerritsen DL, Vernooij-Dassen MJ, Smalbrugge M, Koopmans RT. Comparative validation of proxy-based Montgomery-Asberg depression rating scale and Cornell Scale for Depression in dementia in nursing home residents with dementia. *American Journal of Geriatric Psychiatry* 2012;**20**(11):985–93.

Lichtenberg 1992
Lichtenberg P, Marcopulos B, Steiner D, Tabscott J. Comparison of the Hamilton Depression Rating Scale and the Geriatric Depression Scale: detection of depression in dementia patients. *Psychological Reports* 1992;**70**(2): 515–21.

Lyketsos 1997
Lyketsos CG, Steele C, Baker L, Galik E, Kopunek S, Steinberg M, et al. Major and minor depression in Alzheimer's disease: prevalence and impact. *Journal of Neuropsychiatry and Clinical Neurosciences* 1997;**9**(4): 556–61.

McKeith 2005
McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB consortium. *Neurology* 2005;**65**:1863–72.

McKhann 1984
McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-DRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;**34**(7): 939–44.

McSweeney 2012
McSweeney K, Jeffreys A, Griffith J, Plakiotis C, Kharsas R, O'Connor DW. Specialist mental health consultation for depression in Australian aged care residents with dementia: a cluster randomized trial. *International Journal of Geriatric Psychiatry* 2012;**27**(11):1163–71.

Muller-Thomsen 2005
Muller-Thomsen T, Arlt S, Mann U, Mass R, Ganzer S. Detecting depression in Alzheimer's disease: evaluation of four different scales. *Archives of Clinical Neuropsychology* 2005;**20**(2):271–6.

Naarding 2002
Naarding P, Leentjens AF, van Kooten F, Verhey FR. Disease-specific properties of the Hamilton Rating Scale for Depression in patients with stroke, Alzheimer's dementia, and Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neurosciences* 2002;**14**(3):329–34.

Nelson 2011
Nelson J, Devanand D. A systematic review and meta-analysis of placeo-controlled antidepressant studies in people with depression and dementia. *Journal of the American Geriatrics Society* 2011;**59**(4):577–85.

Olin 2002
Olin J, Katz I, Meyers B, Schneider L, Lebowitz B. Provisional diagnostic criteria for the depression of Alzheimer disease: rationale and background. *The American Journal of Geriatric Psychiatry* 2002;**10**(2):129–41.

Orgeta 2014
Orgeta V, Qazi A, Spector AE, Orrell M. Psychological treatments for depression and anxiety in dementia and mild cognitive impairment. *Cochrane Database*

of Systematic Reviews 2014, Issue 1. [DOI: 10.1002/14651858.CD009125.pub2]

Payne 2002
Payne JL, Sheppard JM, Steinberg M, Warren A, Baker A. Steele C, et al. Incidence, prevalence, and outcomes of depression in residents of a long-term care facility with dementia. *International Journal of Geriatric Psychiatry* 2002; **17**(3):247–53.

Pieper 2011
Pieper MJ, Achterberg WP, Francke AL, van der Steen JT, Scherder EJ, Kovach CR. The implementation of the serial trial intervention for pain and challenging behaviour in advanced dementia patients (STA OP!): a clustered randomized controlled trial. *BMC Geriatrics* 201;**11**(1):12.

Prince 2013
Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and meta-analysis. *Alzheimer's & Dementia* 2013;**9**(1): 63–75.

Regier 1994
Regier DA, Kaelber CT, Roper MT, Rae DS, Sartorius N. The ICD-10 clinical field trial for mental and behavioral disorders: results in Canada and the United States. *American Journal of Psychiatry* 1994;**151**(9):1340–50.

Reisberg 1982
Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. *The American Journal of Psychiatry* 1982;**139**(9): 1136–9.

Reitsma 2005
Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–90.

Review Manager 2014 [Computer program]
The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2015
Riley RD, Ahmed I, Ensor J, Takwoingi Y, Kirkham A, Morris RK, et al. Meta-analysis of test accuracy studies: an exploratory method for investigating the impact of missing thresholds. *Systematic Reviews* 2015;**4**:12. [DOI: 10.1186/2046-4053-4-12]

Roman 1993
Roman GC, Tatemichi TK, Erkinjuntti, T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: diagnostic criteria for research studies. *Neurology* 1993;**34**:250–60.

Rutter 2001
Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001;**20**(19):2865–84.

Schreiner 2003
Schreiner AS, Hayakaqa H, Morimoto T, Kakuma T. Screening for late life depression: cut-off scores for the Geriatric Depression Scale and the Cornell Scale for Depression in Dementia among Japanese subjects. *International Journal of Geriatric Psychiatry* 2003;**18**(6): 498–505.

Seitz 2012
Seitz DP, Brisbin S, Herrmann N, Rapoport MJ, Wilson K, Gill SS, et al. Efficacy and feasibility of nonpharmacological interventions for neuropsychiatric symptoms of dementia in long term care: a systematic review. *Journal of the American Medical Directors Association* 2012;**13**(6):503–6.

Seitz 2013
Seitz DP, Gill SS, Herrmann N, Rapoport MJ, Rines J, Wilson K, et al. Pharmacological treatments for neuropsychiatric symptoms of dementia in long-term care: a systematic review. *International Psychogeriatrics* 2013;**25** (2):185–203.

Sink 2005
Sink KM, Holden KF, Yaffe K. Pharmacological treatment of neuropsychiatric symptoms of dementia: a review of the evidence. *JAMA* 2005;**293**(5):596–608.

Snow 2005
Snow AL, Kunik ME, Molinari VA, Orengo CA, Doody R, Graham DP, et al. Accuracy of self-reported depression in persons with dementia. *Journal of the American Geriatrics Society* 2005;**53**(3):389–96.

Spitzer 1979
Spitzer RL, Forman JB, Nee J. DSM-III field trials: I. Initial interrater diagnostic reliability. *The American Journal of Psychiatry* 1979;**136**(6):815–7.

Starkstein 2005
Starkstein SE, Ingram L, Garau M, Mizrahi R. On the overlap between apathy and depression in dementia. *Journal of Neurology, Neurosurgery & Psychiatry* 2005;**76**(8):1070–4.

Steinberg 2004
Steinberg M, Tschanz JT, Corcoran C, Steffens DC, Norton MC, Lyketsos CG, et al. The persistence of neuropsychiatric symptoms in dementia: the Cache County Study. *International Journal of Geriatric Psychiatry* 2004;**19** (1):19–26.

Sunderland 1988
Sunderland T, Alterman I, Yount D, Hill JL, Tariot PN, Newhouse PA, et al. A new scale for the assessment of depressed mood in demented patients. *The American Journal of Psychiatry* 1988;**145**(8):955–9.

Takwoingi 2013
Takwoingi Y, Leeflang MM, Deeks JJ. Empirical evidence of the importance of comparative studies of diagnostic test accuracy. *Annals of Internal Medicine* 2013;**158**(7):544–54.

Takwoingi 2015
Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with

few studies or sparse data. *Statistical Methods in Medical Research* 2015;**Jun 26**:pii: 0962280215592269.

Teri 1997
Teri L, Logsdon R, Uomoto J, McCurry S. Behavioural treatment of depression in dementia patients: a controlled clinical trial. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 1997;**52**(4):159–66.

Teri 2003
Teri L, Gibbons LE, McCurry SM, Logsdon RG, Buchner DM, Barlow WE, et al. Exercise plus behavioral management in patients with Alzheimer disease: a randomized controlled trial. *JAMA* 2003;**290**(15): 2015–22.

Teri 2005
Teri L, McKenzie G, LaFazia D. Psychosocial treatment of depression in older adults with dementia. *Clinical Psychology: Science and Practice* 2005;**12**(3):303–16.

van Asch 2013
van Asch I, Nuyen J, Veerbeek M, Frijters D, Achterberg W, Pot A. The diagnosis of depression and use of antidepressants in nursing home residents with and without dementia. *International Journal of Geriatric Psychiatry* 2013;**28**(3): 312–8.

van Enst 2014
van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Medical Research Methodology* 2014;**14**(1):70.

Volicer 2011
Volicer L, Frijters D, van der Steen J. Underdiagnosis and undertreatment of depression in nursing home residents. *European Geriatric Medicine* 2011;**2**(6):332–7.

Weiner 2001
Weiner MF, Doody RS, Sairam R, Foster B, Liao TY. Prevalence and incidence of major depressive disorder in Alzheimer's disease: findings from two databases. *Dementia and Geriatric Cognitive Disorders* 2001;**13**(1):8–12.

Whiting 2011a
Whiting PF, Rutjes AQ, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.

Whiting 2011b
Whiting P, Westwood M, Beynon R, Burke M, Sterne JAC, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology* 2011;**64**(6):602–7.

Winter 2011
Winter Y, Korchounov A, Zhukova T, Bertschi N. Depression in elderly patients with Alzheimer dementia or vascular dementia and its influence on their quality of life. *Journal of Neurosciences in Rural Practice* 2011;**2**(1):27–32.

World Health Organization 1992
World Health Organization. International Statistical Classification of Diseases and Related Health Problems. *International Statistical Classification of Diseases and Related Health Problems*. 10. Geneva: World Health Organization, 1992.

* *Indicates the major publication for the study*

APPENDICES

Appendix 1. Characteristics of depression rating scales

Scale	Method of administration	Number of scale items	Scoring of scale items	Range and suggested cut-points	Comments
Cornell Scale for Depression in Dementia	Clinician interview of participant and informant	19	Unable to evaluate 0 = absent 1 = mild to intermittent 2 = severe	0 to 28 ≥ 9 screen positive	Specific to depression in dementia

(*Continued*)

Dementia Mood Assessment Scale	Caregiver reports based on clinician interview	24	0 = normal or mild 6 = severe or significant symptoms on scale item	0 to 144 Cut-points not reported	Contains both mood and cognitive assessments
Nurses' Observation Scale for Geriatric Patients	Caregiver report	30 total, 5 items in mood subscale	0 = all of the time 5 = never	0 to 150 0 to 25 for mood subscale ≥ 10 suggestive of depression	
Geriatric Depression Scale	Self report	30	Yes/No Some items are reverse scored (e.g. Yes can indicate possible sign of depression or absence of depression depending on item)	0 to 30 ≥6, ≥ 8, or ≥ 9 screen positive for depression	Multiple versions available, 30 item, 15 item
Hamilton Depression Rating Scale	Clinician interview	17	0 = absent 2 to 4 = severe or significant symptoms (some items are scored 0 to 2, others 0 to 4)	0 to 52 ≥ 8 considered positive for depression	Multiple versions
Montgomery-Åsberg Depression Scale	Clinician interview	10	0 = absent or normal 6 = persistent or severe symptoms	0 to 60 ≥ 13 screen positive for depression	
Beck Depression Index	Self report	21	0 = symptom absent or no change 3 = severe or significant	0 to 63 ≥ 13 screen positive for depression	
Patient Health Questionnaire	Self report	9	0 = absent 3 = symptom present nearly everyday	0 to 27 ≥ 8 to 11 suggestive of depression	

Appendix 2. MEDLINE search strategy

1. exp Dementia/
2. dement*.mp.
3. alzheimer*.mp.
4. (VaD or VCI or "vascular cognit* impair*").mp.
5. (lewy* adj2 bod*).mp.
6. (LBD or DLB).mp.
7. (FTD or FTLD or frontotemp* or "fronto-temp*").mp.
8. or/1-7
9. (depression or depressed or depressive* or MDD).mp.
10. exp Depression/
11. exp Depressive Disorder, Major/
12. Depressive Disorder/
13. or/9-12
14. (hamilton adj6 depress*).mp.
15. (HDRS or "HDR scale" or HAM-D or HAMD or HAMD-17 or HAMD17).mp.
16. "hamilton scale questionnaire".mp.
17. (hamilton adj3 "17-item").mp.
18. (montgomery adj6 depress*).mp.
19. "montgomery-asberg".mp.
20. MADRS.mp.
21. "montgomery scale".mp.
22. "geriatric depression scale".mp.
23. "geriatric depression scale-short".mp.
24. GDS.mp.
25. GDS-15.mp.
26. "nurs* observation* scale for geriatric patients".mp.
27. NOSGER.mp.
28. "patient* health questionnaire*".mp.
29. (PHQ9 or PHQ-9).mp.
30. (PHQ adj6 depress*).mp.
31. (PHQ2 or PHQ-2).mp.
32. ("Cornell scale" adj4 depress*).mp.
33. CSDD.mp.
34. "Beck depression inventory".mp.
35. BDI.mp.
36. "Beck inventory".mp.
37. "Hospital Anxiety and Depression Scale".mp.
38. HADS.mp.
39. (depress* adj2 score*).mp.
40. (depress* adj3 scale*).mp.
41. (depress* adj3 rating*).mp.
42. or/14-41
43. exp "sensitivity and specificity"/
44. "reproducibility of results"/
45. diagnos*.ti.
46. di.fs.
47. du.fs.
48. sensitivit*.ab.
49. specificit*.ab.
50. (ROC or "receiver operat*").ab.
51. Area under curve/

52. ("Area under curve" or AUC).ab.
53. sROC.ab.
54. accura*.ti,ab.
55. (likelihood adj3 (ratio* or function*)).ab.
56. ((true or false) adj3 (positive* or negative*)).ab.
57. ((positive* or negative* or false or true) adj3 rate*).ti,ab.
58. ("positive predictive value" or PPV).ab.
59. ("negative predictive value" or NPV).ab.
60. or/43-59
61. 8 and 42
62. 8 and 13 and 60
63. 61 or 62
Search narrative:
The MEDLINE search strategy above has been created to optimise sensitivity. The strategy utilises a number of concepts:

Concept A: lines 1 to 7 = health condition/s of interest

Concept B: lines 9 to 12 = target condition being measured by the index test/s/the index test/s

Concept C: lines 14 to 41 = index test/s
Concept D: lines 43 to 59 = methodological filter
Two combinations have been used:
A AND C
A AND B AND D
These have then be OR-ed

Appendix 3. Data extraction form for included studies

BIBLIOGRAPHIC DATA	
Author/Year:	
Title (Citation):	
Country:	
BASIC DEMOGRAPHIC AND CLINICAL DATA	
Participant Recruitment:	
Sampling Procedure:	
Setting:	
Number of Participants:	
Gender:	
Age:	

(*Continued*)

APOE4 carrier (%):	
MMSE:	
Education:	
STUDY DESIGN:	
TARGET CONDITION:	
REFERENCE STANDARD(S)	
INDEX TEST:	
Method of the index test administration/scoring system:	
Threshold(s) of index test in each study (cut-points used to define a positive screen):	
Length of time between administration of index test and reference standard:	
Number of disease positive (D+) and disease negative (D-) participants: Number of test positive (T+) and test negative (T-) participants:	
DATA FOR 2X2 TABLE	
True Positive (TP):	Number of participants with index test+ MDD present
False Positive (FP):	Number of participants with index test+ MDD absent
False Negative (FN):	Number of participants with index test- MDD present
True Negative (TN):	Number of participants with index test- MDD absent
Other relevant information:	Sensitivity/Specificity/PPV/ NPV/LR+/LR-/AUC/Prevalence
Data extracted by:	

Appendix 4. QUADAS-2 Assessment

Domain	Participant selection	Index test	Reference standard	Flow and timing
Description	Describe methods of participant selection: describe included participants (prior testing, presentation, intended use of index test and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any participants who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram); describe the time interval and any interventions between index test(s) and reference standard
Signalling questions (yes/no/unclear)	Was a consecutive or random sample of patients enrolled? Was a case-control design avoided? Did the study avoid inappropriate exclusions?	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it prespecified?	Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?	Was there an appropriate interval between index test(s) and reference standard? Did all participants receive the same reference standard? Were all participants included in the analysis?
Risk of bias: (high/low/unclear)	Could the selection of participants have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
Concerns regarding applicability: (yes/no/unclear)	Are there concerns that the included participants do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	-

Anchoring statements to assist with assessment of risk of bias

Domain 1: Participant selection

Risk of bias: could the selection of participants have introduced bias? (high, low, unclear)

Signalling question 1: Was a consecutive or random sample of patients enrolled?

We will score 'yes' for studies that used either a random or consecutive sampling of individuals with dementia; 'no' if a convenience sample is used; and 'unclear' if the study did not specify the method for participant selection or if the information on this item was not provided.

Signalling question 2: Was a case-control design avoided?

We will exclude studies using a case-control design; we will include only studies rated 'yes' on this question and exclude all studies rated 'no'. We will exclude case-control studies because major depressive disorder is commonly underestimated, and use of a case-control study design may overestimate the diagnostic accuracy of tests by only including individuals with more severe major depressive disorder.

Signalling question 3: Did the study avoid inappropriate exclusions?

We will score this as 'yes' if all individuals with dementia who would potentially be eligible for assessment with depression rating scales are included. We will rate this item as 'no' if only a sample of individuals are selected for assessment with the index and reference test based on some other screening criteria prior to receiving the index or reference test (that is screening positive on a brief assessment before undergoing further assessment with index test and reference standard). We will score this item as 'unclear' if there is insufficient information available to determine a rating.

Applicability: are there concerns that the included participants do not match the review question?

We will rate this item as 'no' if the study examined a population of individuals with dementia using both the index test and reference standard, since this would ensure the individuals were more representative of the target population. We will rate this item as 'yes' if only a subpopulation of the entire study is evaluated or if the study evaluates a population such as a psychiatric inpatient unit or mental health setting where the prevalence of major depressive disorder among individuals with dementia would be expected to be higher than in other settings (for example community or long-term care). We will rate this item as 'unclear' if there is insufficient information to determine the applicablity.

Domain 2: Index test

Risk of bias: could the conduct or interpretation of the index test have introduced bias? (high/low/unclear)

Signalling question 1: Were the index test results interpreted without knowledge of the reference standard?

We will rate this item as 'yes' if the term 'blinded' was used to describe the process by which the results of the reference standard were performed with respect to knowledge of the index depression rating scales. We will also rate this item as 'yes' if it is clear in the study that the individuals who completed the index tests were different from those who completed the reference standards (that is research assistants completing the index tests and psychiatrists completing the index test). We will rate this item as 'no' if the same individual administered the test or if knowledge of the results of the index tests was reported to be known to the individual administering the reference standard. We will rate this item as 'unclear' if there is insufficient information available to complete a rating for this item.

Signalling question 2: Were the index test thresholds prespecified?

We will rate this item as 'yes' if the thresholds for determining major depressive disorder status were prespecified or if multiple thresholds were examined in the study; 'no' if only a single threshold was defined post-hoc based on the results of the study; and 'unclear' if there is insufficient information to make a determination about this item.

Were sufficient data on depression rating scale application given for the test to be repeated in an independent study?.

We will score this as 'yes' if the background of the person administering the index test (for example trained research assistant, physician, psychologist) and the sources of information used for the test (for example participant only or participant and other sources of information such as family members or staff) were reported; 'no' if information on both of these items was not reported; and 'unclear' if there was insufficient information to make a determination.

Applicability: are there concerns that the index test, its conduct, or interpretation differ from the review question? (no, yes, unclear)

We will assign 'no' if the depression rating scales were administered to individuals with dementia at risk for depression and conducted according to standard procedures recommended for the scale. We will assign 'yes' if the depression rating scale is conducted in a fashion that is significantly different from accepted procedures recommended for conducting the test, or 'unclear' if there is insufficient information.

Domain 3: Reference standard

Risk of bias: could the reference standard, its conduct, or its interpretation have introduced bias? (high, low, unclear)

Signalling question 1: Is the reference standard likely to correctly classify the target condition?

We will assign a score of 'yes' if the study used one of the validated reference standards listed in the protocol and if the reference standard was administered by a mental health professional (for example psychiatrist, geriatric psychiatrist, psychologist). We will rate studies as 'no' if a validated reference standard was either not used in the study or if it was administered by people other than mental health professionals. We will rate studies as 'unclear' where there is insufficient evidence to determine this item.

Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

We will assign a score of 'yes' if the study reported that individuals administering the reference test were either blinded to the results of the index test or if the individuals administering the index and reference tests were different individuals. We will assign scores of 'no' if it is clear that the individuals administering the reference standard were aware of the results of the index test or if the same individual administered both the index and reference tests.

Signalling question 3: Was sufficient information on the method of depression assessment given for the assessment to be repeated in an independent study?

We will rate this item as 'yes' if the background of the person administering the reference standard was reported (for example psychiatrist) and the method for obtaining the diagnosis (for example clinical interview) was reported. We will rate this item as 'no' if who administered

the reference standard or what information was utilized to determine the reference standard was not clearly identified. We will assign 'unclear' if there is insufficient information to determine this item.

Applicability: are there concerns that the target condition as defined by the reference standard does not match the review question? (no, yes, unclear)

If the reference standard was applied to an unspecified sample of individuals with dementia, we will assign a score of 'no'; if the reference standard was only applied to subgroup of the study sample (for example only those with known depressive symptoms), we will rate this as 'yes'.

Domain 4: Patient flow and timing (Figure 1)

Risk of bias: could the patient flow have introduced bias? (high, low, unclear)

Signalling question 1: Was there an appropriate interval between the index test and reference standard?

Depression can resolve or remit over time, although it tends to be relatively stable over short periods of time. We will rate this item as 'yes' if the index and reference tests were completed within a two-week period, 'no' if a period of greater than two weeks elapsed between administration of the index and reference tests, and 'unclear' if the interval between the the index test and reference standard was not specified.

Signalling question 2: Did all participants receive the same reference standard?

We will rate this item as 'yes' if all individuals who received the index test also received the reference standard and 'no' if only a sample of individuals that received the index test received the reference standard (for example only those who screened positive on the index test later received the reference standard). We will assign 'unclear' if the proportion of individuals who received both the index and reference standard was not clear.

Signalling question 3: Were all participants included in the final analysis?

We anticipate that dropouts will be minimal in the cross-sectional studies included in our review. We will assign a score of 'yes' if 90% or greater of individuals who received the index test later had the reference standard, 'no' if this proportion is less than 90%, and 'unclear' where there is insufficient information to make a determination.

CONTRIBUTIONS OF AUTHORS

All review authors contributed to the development of the research protocol and have approved the final version for submission.

DECLARATIONS OF INTEREST

Dallas Seitz served on an advisory board for Eli Lilly in 2013.

Julia G Kirkham - none known

Yemisi Takwoingi - none known

Terry J Quinn - none known

Mark Rapoport - none known

Krista L Lanctôt - none known

Colleen J Maxwell - none known

Nathan Herrmann - none known

Sudeep S Gill - none known

Paula Rochon - none known

SOURCES OF SUPPORT

Internal sources

- Canadian Institutes of Health Research, Canada.

External sources

- Canadian Institutes of Health Research, Canada.

This project is funded by a Canadian Institutes of Health Research Knowledge Synthesis Grant KRS#141000

- NIHR, UK.

This protocol was supported by the National Institute for Health Research, via Cochrane Infrastructure funding to the Cochrane Dementia and Cognitive Improvement group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health